

What Is Claimed Is:

- 1 1. A method for facilitating use of a collation element that supports a
2 large number of characters, comprising:
3 receiving the collation element;
4 reading a primary weight value from a primary weight field within the
5 collation element;
6 if the primary weight value falls within a reserved set of values, reading an
7 additional portion of the primary weight value from a secondary weight field and a
8 tertiary weight field within the collation element; and
9 if the primary weight value is not within the reserved set of values,
10 reading a secondary weight value from the secondary
11 weight field within the collation element, and
12 reading a tertiary weight value from the tertiary weight field
13 within the collation element.
- 1 2. The method of claim 1, wherein if the primary weight value falls
2 within a reserved set of values, the method additionally comprises:
3 setting the secondary weight value to a secondary default value; and
4 setting the tertiary weight value to a tertiary default value.
- 1 3. The method of claim 1, wherein the collation element adheres to a
2 structure specified in Unicode Technical Report No. 10.
- 1 4. The method of claim 1,
2 wherein the primary weight value identifies a character;

3 wherein the secondary weight value can specify an accent on the character;
4 and
5 wherein the tertiary weight value can specify case information for the
6 character.

1 5. The method of claim 1, wherein the collation element is four bytes
2 in size, of which the primary weight field is two bytes, the secondary weight field
3 is one byte and the tertiary weight field is one byte, unless a value in the primary
4 weight field belongs to the reserved set of values, in which case the primary
5 weight field takes up all four bytes of the collation element.

1 6. The method of claim 5, wherein the reserved set of values for the
2 primary weight value includes hexadecimal values 0xFFFF0-0xFFFF.

1 7. The method of claim 1, wherein the collation element is taken from
2 a collation weight table that is used to map characters to collation weights in order
3 to establish an ordering between strings of characters.

1 8. The method of claim 7, further comprising constructing a sorting
2 key for a string by:
3 reading each character in the string;
4 looking up a corresponding collation element for each character from the
5 collation weight table; and
6 adding the corresponding collation element for each character to the
7 sorting key.

1 9. The method of claim 8,

2 wherein the sorting key is associated with a record within a database; and
3 wherein the sorting key is used to construct a linguistic index for the
4 database.

1 10. A computer-readable storage medium storing instructions that
2 when executed by a computer cause the computer to perform a method for
3 facilitating use of a collation element that supports a large number of characters,
4 the method comprising:
5 receiving the collation element;
6 reading a primary weight value from a primary weight field within the
7 collation element;
8 if the primary weight value falls within a reserved set of values, reading an
9 additional portion of the primary weight value from a secondary weight field and a
10 tertiary weight field within the collation element; and
11 if the primary weight value is not within the reserved set of values,
12 reading a secondary weight value from the secondary
13 weight field within the collation element, and
14 reading a tertiary weight value from the tertiary weight field
15 within the collation element.

1 11. The computer-readable storage medium of claim 10, wherein if the
2 primary weight value falls within a reserved set of values, the method additionally
3 comprises:
4 setting the secondary weight value to a secondary default value; and
5 setting the tertiary weight value to a tertiary default value.

1 12. The computer-readable storage medium of claim 10, wherein the
2 collation element adheres to a structure specified in Unicode Technical Report
3 No. 10.

1 13. The computer-readable storage medium of claim 10,
2 wherein the primary weight value identifies a character;
3 wherein the secondary weight value can specify an accent on the character;
4 and
5 wherein the tertiary weight value can specify case information for the
6 character.

1 14. The computer-readable storage medium of claim 10, wherein the
2 collation element is four bytes in size, of which the primary weight field is two
3 bytes, the secondary weight field is one byte and the tertiary weight field is one
4 byte, unless a value in the primary weight field belongs to the reserved set of
5 values, in which case the primary weight field takes up all four bytes of the
6 collation element.

1 15. The computer-readable storage medium of claim 14, wherein the
2 reserved set of values for the primary weight value includes hexadecimal values
3 0xFFFF0-0xFFFF.

1 16. The computer-readable storage medium of claim 10, wherein the
2 collation element is taken from a collation weight table that is used to map
3 characters to collation weights in order to establish an ordering between strings of
4 characters.

1 17. The computer-readable storage medium of claim 16, wherein the
2 method further comprises constructing a sorting key for a string by:
3 reading each character in the string;
4 looking up a corresponding collation element for each character from the
5 collation weight table; and
6 adding the corresponding collation element for each character to the
7 sorting key.

1 18. The computer-readable storage medium of claim 17,
2 wherein the sorting key is associated with a record within a database; and
3 wherein the sorting key is used to construct a linguistic index for the
4 database.

1 19. An apparatus that facilitates use of a collation element that
2 supports a large number of characters, comprising:
3 an assignment mechanism that is configured to read a primary weight
4 value from a primary weight field within the collation element;
5 wherein if the primary weight value falls within a reserved set of values,
6 the assignment mechanism is configured to read an additional portion of the
7 primary weight value from a secondary weight field and a tertiary weight field
8 within the collation element; and
9 wherein if the primary weight value is not within the reserved set of
10 values, the assignment mechanism is configured to,
11 read a secondary weight value from the secondary weight
12 field within the collation element, and to
13 read a tertiary weight value from the tertiary weight field
14 within the collation element.

1 20. The apparatus of claim 19, wherein if the primary weight value
2 falls within the reserved set of values, the assignment mechanism is configured to:
3 set the secondary weight value to a secondary default value; and to
4 set the tertiary weight value to a tertiary default value.

1 21. The apparatus of claim 19, wherein the collation element adheres
2 to a structure specified in Unicode Technical Report No. 10.

1 22. The apparatus of claim 19,
2 wherein the primary weight value identifies a character;
3 wherein the secondary weight value can specify an accent on the character;
4 and
5 wherein the tertiary weight value can specify case information for the
6 character.

1 23. The apparatus of claim 19, wherein the collation element is four
2 bytes in size, of which the primary weight field is two bytes, the secondary weight
3 field is one byte and the tertiary weight field is one byte, unless a value in the
4 primary weight field belongs to the reserved set of values, in which case the
5 primary weight field takes up all four bytes of the collation element.

1 24. The apparatus of claim 23, wherein the reserved set of values for
2 the primary weight value includes hexadecimal values 0xFFFF0-0xFFFF.

1 25. The apparatus of claim 19, wherein the collation element is taken
2 from a collation weight table that is used to map characters to collation weights in
3 order to establish an ordering between strings of characters.

1 26. The apparatus of claim 25, further comprising a key construction
2 mechanism for constructing a sorting key for a string, wherein the key
3 construction mechanism is configured to:
4 read each character in the string;
5 lookup a corresponding collation element for each character from the
6 collation weight table; and to
7 add the corresponding collation element for each character to the sorting
8 key.

1 27. The apparatus of claim 26,
2 wherein the sorting key is associated with a record within a database; and
3 wherein the sorting key is used to construct a linguistic index for the
4 database.